

CHAPTER III

RESEARCH METHOD

A. Research Design

This type of research is quasi-experimental with a quantitative approach. The quantitative approach is a research method based on the postpositivist worldview, which emphasizes cause-and-effect relationships, the reduction of variables into testable hypotheses, and the use of standardized instruments to collect numerical data that is then statistically analyzed. This approach aims to test theories and produce generalizations based on objectively measurable data (Creswell, 2013). Quantitative methods are defined as research methods based on positivist philosophy, used to study specific populations or samples through standardized research instruments. This method analyzes data objectively, measurably, rationally, and systematically, and it continues to maintain its scientific characteristics even as it adapts to technological developments in the Industrial Revolution 4.0 and 5.0 eras. In these eras, research processes have become integrated with digital systems and modern statistical applications, while still using the classic formulas as their foundation (Susanto & Jaya, 2023). The quasi-experimental design allows researchers to conduct causal analysis even without random assignment, as long as threats to validity can be identified and controlled (Campbell & Riecken, 1968). The

study aims to measure the effect of implementing the Project-Based Learning (PjBL) method on improving students' speaking skills at SMAN Keberbakatan Olahraga Kota Bengkulu. The research employs a "Non-equivalent Control Group Design," where two groups of students are selected: one as the experimental group using the PjBL method and the other as the control group using conventional teaching methods.

Table 3. 1 The Experiment and Control of the Research

Experiment	O ₁	x	O ₂
Control	O ₃	-	O ₄

(Source: Sugiyono, 2011)

Description:

- E : Experimental group
- K : Control group
- X : Application Media
- O₁ : Pre test in experimental class
- O₂ : Post test on the experimental class
- O₃ : Pre test in control class
- O₄ : Post test on control class

B. Place and Time of the Research

This research was conducted on 11th grade students of SMAN Keberbakatan Olahraga in Bengkulu City for the 2025/2026 academic year, located in Bengkulu City. The 11th grade students were chosen as the research subjects. The population in this study consisted of 60 students. The researcher

selected this class because the students had good English speaking skills. The research was conducted in August 13, 2025. The researcher gave a pre-test in the form of a dialogue to the students of class XI A as the experimental class in the first meeting, and then the students' dialogues were recorded for analysis. In the following meetings, the researcher carried out treatments using the Project-Based Learning method four times in class XI A. After the treatments were finished, the researcher conducted a post-test using the same dialogue instrument as in the pre-test, and the results were also recorded. At the same time, the researcher also gave the same pre-test in the form of a dialogue to the students of class XI B as the control class, without giving any treatment. The control class only followed the usual learning activities. Afterwards, after the learning period ended, the researcher gave a post-test in the form of the same dialogue as in the pre-test to class XI B, and the results were also recorded for comparative analysis with the experimental class.

C. Population and Sample

1. Population

The population is the larger group to which the researcher intends to generalize the results; it includes all individuals or elements that share the characteristics being studied (*Casteel & Bridier, 2021*). In this study, the selected population is 11th-grade students, as at this level they have

a more mature understanding of the learning material and demonstrate cognitive development that is appropriate to support effective data collection.

Table 3. 2 The Populations of the Research

No.	Class	Total Number of Students
1.	XI.1	20
2.	XI.2	20
3.	XI.3	20
	Total	60

(Source: SMAN Keberbakatan Olahraga Kota Bengkulu, 2025)

2. Sample

The sample for this study is selected using purposive sampling. Purposive sampling is an iterative process of selecting research subjects based on their particular knowledge or experience with the phenomenon under investigation, rather than using a predetermined sampling frame. (Robinson, 2024, p. 5645–5647). The researcher selected one class from the three 11th-grade classes at SMAN Keberbakatan Olahraga Kota Bengkulu as the sample.

The selected class consists of 40 students, based on the following considerations:

- a. The alignment of the class schedule with the research needs.
- b. The homogeneity of students' speaking skills as determined through discussions with the English teacher.

- c. Accessibility and ease of coordination for conducting the study.

Table 3. 3 The Sample of the Research

No.	Class	Total Number of Students
1.	XI.1 (Experiment Class)	20
2.	XI.2 (Control Class)	20
	Total	40

(Source: SMAN Keberbakatan Olahraga Kota Bengkulu, 2025)

D. Definitions of Operational Variable

According to Creswell (2018), an operational definition refers to a specific way of defining a variable so that it can be measured quantitatively, and this is important to ensure that the terms used in the study maintain consistent and fixed meanings throughout the research.

1. Project-Based Learning

Project Based Learning (PjBL) is a learning model that engages students in in-depth investigation of real-world problems to produce meaningful products. This approach encourages collaboration, independence, and the development of critical thinking and problem-solving skills (Thomas, 2000).

2. Speaking Skill

According to Brown and Lee (1995), speaking skills are an interactive process of constructing meaning that involves the simultaneous production, reception, and processing of information. They emphasize that teaching

speaking skills should encourage interaction, use authentic contexts, and provide constructive feedback so that students are motivated and able to communicate effectively.

E. Technique of Data Collecting

1. Test

The following action were taken by the research to gather the data:

a. Pre-Test

The first step was to conduct a pre-test using a speaking test instrument to measure students' initial abilities in fluency, pronunciation, grammar accuracy, vocabulary, and comprehensibility. This pre-test result was the basic data to be compared with the post-test result after the implementation of PjBL.

b. Treatment

This study applied the Project-Based Learning (PjBL) method to the experimental group for four sessions. Each session integrated collaboration, problem solving, and English usage through projects in the form of dialogues related to students' daily lives, such as: "What do you think the new cafeteria menu?", "What's your opinion on the new action movie?", "Do you have any idea about where we should go for the class trip?", "What is your point of view on the new math teacher?", "Do you have any opinion on painting as a hobby?", "What do you think of using tablet in the

classroom?", "How do you feel about the new spicy ramen at the canteen?", "What about the new song by your favorite band?", "What is your reaction to the ending of the novel?", and "What's your opinion on banning plastic bags?". The main purpose of these activities was to improve students' speaking skills, including fluency, pronunciation, grammar, vocabulary, and comprehension. It was expected that the post-test results would show significant improvements compared to the pre-test.

c. Post-test

Post-test was conducted using the same instrument as the pre-test. The results of this post-test were compared with the pre-test to measure the extent to which PjBL influenced the improvement of students' speaking skills.

2. Try Out

The try-out was given in the try-out class to evaluate the test items in order to find out whether the test was valid or not before being used for the pre-test and post-test. The trial was given to XI i students of SMAN 3 Kota Bengkulu. The reason the researcher chose class XI i of SMAN 3 Kota Bengkulu as the location was because it was at the same level as class XI at SMAN Keberbakatan Olahraga Kota Bengkulu.

The researcher conducted a try-out on July 25, 2025, at SMAN 3 Kota Bengkulu as an initial stage to validate the test items that would be used in the study. The try-out involved 18 students as respondents and consisted of 20 test items. Based on the analysis results, 14 items were found to be valid and 6 items were invalid. The results of this try-out served as the basis for determining the items to be used in the pre-test and post-test stages.

Table 3. 4 The Grid of Speaking Skill Test Instrument

No.	Aspect of Variable	Indicator	Items
1.	Pronunciation	Students pronounce words clearly and correctly in giving opinion dialogues	Dialogue 1,2,6,7,10,11,13,14,18, and 20
2.	Vocabulary	Students use appropriate and context-relevant vocabulary in their responses	Dialogue 2,3,5,6,10,11,12,14,15,16,18, and 20
3.	Grammar	Students apply correct grammar	Dialogue 1,4,6,8,9,11,12,14,15,17,18, and 20

		structures in speaking	
4.	Fluency	Students speak smoothly with minimal pauses or hesitation	Dialogue 3,4,6,7,9, 11, 12, 13, 14, 15, 18, and 20
5.	Comprehension	Students understand and respond appropriately to opinions given	Dialogue 2,6,9, 11, 12, 13, 14, 15, 16, 18, and 20

(Source; Managed, 2025)

1. Project-Based Learning

Project-based learning is a teaching method in which students gain knowledge and skills by working for an extended period of time to investigate and respond to a complex question, problem, or challenge.

(Solomon, 2003, p. 20)

2. Speaking Skill

Speaking skills can be assessed through aspects of pronunciation, grammar, vocabulary, fluency, and comprehension to reflect students' overall communication ability in the classroom (Heaton, 1979). According to Bashir, Azeem, and Dogar (2011), English speaking skill is a productive ability that involves not only the pronunciation of words, but also the mastery of grammar, vocabulary, and

an understanding of appropriate social and cultural contexts to create effective communication.

3. Documentation

Documentation is an important part of the research process because it helps ensure the validity, transparency, and ethics of data reporting (Creswell, 2013). The documentation of this research include:

- a. The number of students
- b. The number of students' class
- c. The students' speaking score
- d. The students' speaking video/record
- e. Photo during the research.

F. Data Analysis Technique

1. Validity of the Test

Validity is an important aspect in research that refers to the extent to which a method truly measures what it is intended to measure. According to Ahmed and Ishtiaq (2021), "validity is defined as to how accurately a methodology measures a variable that it intends to measure." This means that validity emphasizes the accuracy of a measuring instrument in representing the variable being studied. In this context, a data collection instrument or method is considered valid if the results obtained genuinely reflect the actual conditions or characteristics of the research subject. In other words, the higher the level of

validity of a method, the greater the likelihood that it will produce data that corresponds to reality and the research objectives. According to Amirrudin, Nasution, and Supahar (2021), validity refers to the extent to which an instrument is capable of measuring what it is intended to measure. Validity reflects the accuracy of the measurement in relation to the targeted construct or concept. An instrument cannot be considered valid if it is not reliable, as inconsistent results cannot serve as a sound basis for accurate assessment. In other words, reliability is a prerequisite for validity, although a reliable instrument is not necessarily valid.

Table 3. 5 Test Instrument Validity Criteria

No.	Cronbach' Alpha	Validity Level
1.	0,81 - 1,00	Highly
2.	0,61 - 0,80	High
3.	0,41 - 0,60	Fairly
4.	0,21 - 0,40	Low
5.	0,00 - 0,20	Very Low

(Source: Widodo at al, 2023)

Table 3. 6 The Grid of Speaking Skill Test Instrument Validity

No.	Aspect of Variable	Indicator	Item Number
1	Pronunciation	Students pronounce words clearly and correctly in giving opinion dialogues.	2, 5, 6, 7, 8, 9, 11, 14, 18, 20

2	Vocabulary	Students use appropriate and context-relevant vocabulary in their responses.	2, 4, 5, 6, 8, 9, 11, 13, 14, 15, 18, 19, 20
3	Grammar	Students apply correct grammar structures in speaking.	2, 4, 5, 6, 9, 11, 13, 14, 15, 17, 18, 19
4	Fluency	Students speak smoothly with minimal pauses or hesitation.	2, 4, 5, 6, 8, 9, 11, 13, 14, 15, 17, 19
5	Comprehension	Students understand and respond appropriately to opinions given.	4, 6, 9, 11, 13, 14, 15, 17, 18, 19, 20

(Source: Managed, 2025)

Based on the SPSS analysis of the test instrument consisting of 20 dialogues, a validity test was conducted using the item-total correlation (Corrected Item-Total Correlation). The validity criterion used was $r\text{-count} > r\text{-table}$, with the $r\text{-table}$ value for $N = 18$ being 0.468. **Table 3. 7 Summary of Validity Test Results**

Summary of Validity Test Results			
No Dialogue	R Count	r table	Status
1	0,4392	0,4683	Invalid
2	0,5558		Valid
3	0,4188		Invalid
4	0,5304		Valid
5	0,5005		Valid
6	0,4641		Invalid
7	0,4873		Valid
8	0,6648		Valid
9	0,5787		Valid
10	0,3660		Invalid
11	0,5976		Valid
12	0,3071		Invalid

13	0,6614		Valid
14	0,6142		Valid
15	0,5123		Valid
16	0,0945		Invalid
17	0,7087		Valid
18	0,5787		Valid
19	0,6851		Valid
20	0,5976		Valid

Based on the calculation results, it was found that 14 dialogue items were declared valid and 6 items were invalid. The valid items had an r-count greater than the r-table, while the invalid items had an r-count smaller or even negative. From the result of analysis, it was found 14 dialogue items were declared valid with the following indicators:

Table 3. 8 Indicators of Dialogue Validity Test

No.	Dialogue	Test Indicator
1.	Lina: What's your opinion on the new action movie? Fahmi: _____ that it was very entertaining.	Grammar, Vocabulary, Fluency
2.	Nadia: What is your point of view on the new math teacher? Bayu: _____ she explains things very clearly.	Grammar, Pronunciation, Comprehension
3.	Arif: Do you have any opinion on painting as a hobby? Tari: _____, I find it relaxing and creative.	Vocabulary, Fluency, Grammar
4.	Bella: How do you feel about the new spicy ramen at the canteen? Rino: _____ it's delicious.	Vocabulary, Pronunciation, Fluency

5.	Fani: What about the new song by your favorite band? Jodi: _____, believe that it's their best song so far.	Vocabulary, Grammar, Fluency
6.	Tina: What is your reaction to the ending of the novel? Kevin: _____ that it was the perfect twist!	Grammar, Vocabulary, Comprehension
7.	Rina: What do you think about online learning? Dio: _____ it's very convenient and flexible.	Grammar, Fluency, Comprehension
8.	Alya: What is your point of view on reading e-books? John: _____ they're practical and easy to carry anywhere.	Vocabulary, Grammar, Pronunciation
9.	Dian: Do you have any opinion about school uniforms? Fajar: _____ uniforms create a sense of equality.	Grammar, Comprehension, Vocabulary
10.	Lusi: What's your reaction to the new school regulations? Ivan: _____ they're too strict, especially the phone policy.	Vocabulary, Grammar, Fluency
11.	Tari: What's your opinion on waking up early for school? Rama: _____ helps create discipline.	Grammar, Fluency, Pronunciation
12.	Yola: How do you feel about learning a second language? Fikri: _____ it's a valuable skill in today's world.	Vocabulary, Comprehension, Fluency
13.	Nina: What about joining extracurricular clubs?	Grammar, Vocabulary, Fluency

	Dhani: _____ they build character and confidence.	
14.	Citra: What's your opinion on watching movies for learning English? Rafa: _____, it makes learning fun and natural.	Vocabulary, Fluency, Comprehension

2. Reliability of the Study

Reliability in research is defined as the consistency of a method in performing measurements. According to Ahmed and Ishtiaq (2021), "reliability is defined as the consistency of a method in measuring something." In other words, a measuring instrument is considered reliable if it can consistently produce the same results when used under similar conditions. This consistency is crucial to ensure that the data obtained is trustworthy and can be reproduced in repeated measurements. According to Amirrudin, Nasution, and Supahar (2021), reliability is the ability of an instrument to produce consistent and stable data over time. A reliable instrument will yield similar results even when used under different conditions or repeated several times. One common method used to measure reliability is Cronbach's Alpha, which assesses internal consistency by examining the extent to which the items in an instrument are correlated with one another. The higher the alpha value, the greater the reliability of the instrument.

Table 3. 9 Test Instrument Reliability Criteria

No.	Cronbach' Alpha	Reliability Level
1.	0,00 – 0,20	Lack of Reliable
2.	0,20 – 0,40	Somewhat Reliable
3.	0,40 – 0,60	Fairly Reliable
4.	0,60 – 0,80	Reliable
5.	0,80 – 1,00	Highly Reliable

(Source: Nuzulia, 2019)

Table 3. 10 The Table of Reliability

Reliability

Case Processing Summary

		N	%
Cases	Valid	18	100.0
	Excluded ^a	0	.0
	Total	18	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.865	14

The reliability test using Cronbach's Alpha showed a value of $\alpha = 0.865$. This value is higher than the minimum criterion of 0.60, indicating that the instrument has highly reliable. This means the instrument possesses good internal consistency and can be trusted to measure students' abilities in a stable manner.

3. Difficulty Level Test

According to Purba, Fadhilaturrahmi, Purba, and Siahaan (2021), the difficulty index test is an analysis used to determine the level of difficulty of a test item based on the proportion of test-takers who answer it correctly. According to Haladyna (2004), the purpose of testing item difficulty is to find out how many students give the correct answer. This can be achieved by knowing how many students work on or choose the correct answer.

Difficulty index:

$$P = \left(\frac{Rh}{Nh} + \frac{Ri}{Ni} \right) \times 100\%$$

Description:

P: Difficulty level in percent

Nh: Number of test takers in the highest score group

Rh: Number of correct answers in the highest score group

Ni: Number of test takers in the low score group

Ri: Number of correct answers in the low score group

$$P = \left(\frac{265}{5} + \frac{80}{5} \right) \times 100\%$$

$$P = 69\%$$

Table 3. 11 Interpretation of Item Difficulty Test

Score P	Interpretation
0.00-0.40	Difficult
0.41-0.70	Medium
0.71-1.00	Easy

(Source: Haladyna, 2004)

The calculation shows that the difficulty index (P) is 69%, which falls in the range of 0.41–0.70, meaning it is in the medium category. A medium-level question is considered ideal because it gives a balance between easy and difficult questions. If a question is too easy, almost all students can answer it correctly, so it cannot show real differences in ability. On the other hand, if a question is too difficult, most students will answer it incorrectly, which also makes it hard to see differences in ability. A medium-level question allows some students to answer correctly while others may answer incorrectly, making it more effective in measuring various levels of ability. In addition, questions in the medium category usually have good discrimination power, meaning they can distinguish between students with high ability and those with lower ability. Therefore, this question can be considered appropriate and ideal to use in a test or research, as it contributes to the overall quality of the instrument.

4. Discriminating Power Test

Purba et al. (2021) explain that the discrimination index test aims to determine a question's ability to distinguish between high- and low-ability test-takers. The following is the formula for computing discriminating power as suggested by Gronlund (1993).

$$DP = \frac{U-L}{N}$$

Description:

DP = Item discrimination

U = Correct responses from the upper group.

L = Correct responses from the lower group.

N = Number of students per group (27%).

$$DP = \frac{265-80}{10}$$

$$DP = 18,5\%$$

Table 3. 12 Discriminating Power

<i>Discriminating Power</i>	<i>Interpretation of Result</i>
$0,71 \leq DP \leq 1,00$	Very good
$0,41 \leq DP \leq 0,70$	Good
$If +0,21 < DP < 0,40$	Sufficiently qualified
$If 0 < DP \leq 0,20$	Weak

(Source: Desman et al, 2025)

Based on the calculation results, the discriminating power (DP) was found to be 18.5%, which is equivalent to 0.185. When compared to the criteria in Table 3.13, this value falls within the range of 0–0.20, which is categorized as weak. This indicates that the test item being analyzed has a reasonable level of discriminating power, although it is not yet optimal. In other words, the item is still able to distinguish between students with high and low ability, but its effectiveness is relatively limited. Therefore, test items in this category can still be used in research instruments.

5. Normality Test

Normality testing is a statistical procedure used to assess whether the obtained data come from a normally

distributed population, which is commonly assumed in various parametric tests such as the t-test and linear regression" (Tsagris & Pandis, 2021). The Shapiro-Wilk test is used to determine whether a dataset comes from a normal distribution, with the null hypothesis stating that the data are normally distributed. If the p-value is smaller than the predetermined significance level, the null hypothesis is rejected (Khatun, 2021).

6. Homogeneity Test

The Homogeneity Test is used to determine whether several population variants are the same or not. This test was carried out as an opening to the Independent Sample t Test and Anova analysis. The assumption underlying analysis of variance (Anova) is that population variances are the same. The equality of two variances test is used to test whether the data distribution is homogeneous or not, namely by comparing the two variances. If two or more groups of data have the same variance, then the Homogeneity test does not need to be carried out again because the data is considered homogeneous. The homogeneity test can be carried out if the data group is in a normal distribution. The homogeneity test is carried out to show that differences that occur in parametric statistical tests (eg test, anova, anacova) do occur due to differences between groups, not due to differences within groups. The homogeneity of variance test is very necessary before comparing two or more groups, so that the

differences are not caused by differences in basic data (not homogeneity of the groups being compared).

7. T-Test

Paired sample t-test, this test is used to see if there is a significant increase between the pre test and post test. according to Pallant (2020), paired sample t-test is used to compare the means of two mutually bound groups, such as testing the difference in scores before and after intervention.

8. F-Test

The F-test is a statistical test used to examine joint linear hypotheses in a linear model by comparing the variation explained by the model with the error variation. The power of the F-test is influenced by several main factors, especially the noncentrality parameter and the degrees of freedom, which determine the ability of the test to reject the null hypothesis when the alternative hypothesis is true. Griffiths and Hill (2022) explain that adding true restrictions to the null hypothesis does not reduce the noncentrality parameter of the F statistic, but it can increase or decrease the power of the test depending on the effect of reducing the numerator degrees of freedom. Therefore, the F-test is not only a tool for testing the significance of joint linear hypotheses, but its power is also strongly affected by the structure of the hypothesis and the characteristics of the model used (Griffiths & Hill, 2022).

9. The Effect Size

This test measures the magnitude of the influence or difference observed in the research. This is very important because it reveals the results regarding how large or strong the effect is on the research. Nordahl-Hansen et al emphasize that the interpretation of effect size is more important than the p-value because it shows the strength of the relationship between variables (Nordahl-Hansen et al., 2024). Effect size is commonly used to measure the average difference between two groups and Pearson's r , which measures how strong the relationship is between two variables. Here are the categories of effect size values that are often used to see how significant the influence is between two variables in research:

Table 3. 13 Value of the Effect Size

No	Value	Effect Size
1.	0.01-0.19	Very Small
2.	0.20-0.49	Small
3.	0.50-0.79	Medium
4.	0.80-1.19	Large
5.	1.20-1.99	Very Large
6.	≥ 2.00	Huge Effect

(Source: Cohen, 2013)

10. Statistical Analysis

Researchers want to find out whether the project-based learning method has an influence on students' speaking skills. The authors also calculated data from the experimental and control classes using the following T-test formula.

- a. The null hypothesis (H_0) and the alternative hypothesis (H_a) are accepted, significant effect of speaking skills method on students'.
- b. The null hypothesis (H_0) is rejected, while the alternative hypothesis (H_a) is accepted, indicating that the influence of project-based learning on students' speaking skills is significant.

